Landsat Science Team

Landsat Product Updates

Level-1 Collection Plans and Issues

12 January 2016

Brian Sauer
USGS EROS
bsauer@usgs.gov

U.S. Department of the Interior
U.S. Geological Survey

Product Improvement & Collection Status

Agenda

- Product improvements to be included in collection 1
- Collection definition (John)
- Collection identification (Chris)
- Collection data management plan (Brian, Kristi)
- Collection management control (Brian)

USGS

Level-1 Product Roadmap Status
(1 of 2)

Improvement	Instrument Data Affected	Description	Collection 1
Quality Band	OLI/TIRS	Updates to include fmask Output	Y
	TM/ETM+	Quality band similar to Landsat 8; utilize fmask for cloud detection	Y
	MSS	Quality band similar to Landsat 8; utilize cubist (existing) algorithm for cloud detection	N
Ground Control Library Updates	TM/ETM+	Phase I - 177 Path/Row combinations improved; GCP water mask	Y
	OLI/TIRS	Phase I - 177 Path/Row combinations improved; GCP water mask	Y
	TM/ETM+, OLI/TIRS	Phase II - Low-latitude areas	Y
	TM/ETM+, OLI/TIRS	Phase III - High-latitude areas	Y
TOA Reflectance Angle Coefficients File (Enhanced Metadata)	OLI/TIRS	Scene-specific per-pixel solar azimuth and sensor viewing angle coefficients (enhanced metadata) to allow users to convert to reflectance	Y
	TM/ETM+	Scene-specific per-pixel solar azimuth and scan angle coefficients (enhanced metadata) to allow users to convert to reflectance	Y

Level-1 Product Roadmap
(2 of 2)

Improvement	Instrument Data Affected	Description	Collection 1
Level 1 data format study	All	More flexible alternatives to gzip'd L1T, more condusive to data delivery services	N
Product 'versioning' updates	All	Improve data so users can more readily determine changes in the product. Implemented through collections and processing date in file name.	Y
Land-based cloud cover score	OLI/TIRS	Calculate CCA on land-only using output from fmask - provide in metadata / user search	Y
	TM/ETM+	Calculate CCA on land-only using output from fmask - provide in metadata / user search	Y
	MSS	Calculate CCA on land-only using output from cubist - provide in metadata / user search	N
Cirrus CCA Improvements	OLI/TIRS	Update cirrus CCA to reduce false positives due to high-terrain areas being visible in the cirrus band.	Y
Landsat DEM Improvement	All	Augment or replace existing GLS2000 DEM	N

Collection 1 Pixel Metadata

Bit	Value	Cumulative Sum	Description – OLI	Description – TM/ETM+	Description – MSS	8-Bit Browse
0	1	1	Designated Fill	Designated Fill	Designated Fill	Designated Fill
1	2	3	Terrain Occlusion	Dropped Pixel	Dropped Pixel	Occluded/Dropped
2	4	7	Radiometric Saturation	Radiometric Saturation	Radiometric Saturation	Saturation
3	8	15				Cloud
4	16	31	Cloud	Cloud	Cloud	Shadow
5	32	63	Cloud Confidence	Cloud Confidence	Cloud Confidence	Snow/Ice
6	64	127				Cirrus
7	128	255	Cloud Shadow	Cloud Shadow		
8	256	511				
9	512	1023	Snow/Ice	Snow/Ice		
10	1024	2047				
11	2048	4095	Cirrus			
12	4096	8191				
13	8192	16383				
14	16384	32767				
15	32786	65553				

USGS

Collection Definition Document Review

- John with input from cal/val, science, and others created a draft Level-1 Product Collection Definition Document
- You are invited to review this document
  - Highlights from the document are included within this slide deck
  - Please provide comments back in any form
    - Prefer track changes on word document or simple email outlining suggestions for improvement of the doc

Landsat Level-1 Product Collection Definition

Summary

The Land Satellites Data System (LSDS) project is proposing to implement Landsat Level-1 data collections of prescribed radiometric and geometric quality. The radiometric performance of the different Landsat sensors can only be characterized and quantified at the instrument level in terms of absolute radiometric uncertainty and temporal uncertainty (stability). The geometric quality of the Landsat Level-1 products, however, can be quantified on a per scene basis for each instrument in terms of geodetic accuracy relative to ground control. The geodetic accuracy of the Level-1 products becomes the discriminating factor in determining whether Level-1 products satisfy the criteria for being in a collection, i.e. the data are "stackable" to enable multispectral time series analysis.

The radiometric uncertainty for Landsat 8 operational land imager (OLI) data is based on top of atmosphere reflectance measured over pseudo-invariant calibration sites (PICS). The trend in OLI radiometric uncertainty is better than .3% and it is better than 2% for the Landsat 7 enhanced thematic mapper plus (ETM+) and Landsat 5 thematic mapper (TM) instruments. The geodetic accuracy of the Level-1 products vary by sensor, the source data type, the quality of the payload correction data (PCD), and the level to which the data have been processed: precision and terrain correction (L1T); systematic and terrain correction (L1GT); and systematic correction (L1G). The older TM-A PCD are of lower quality, and for some of the TM data being repatriated through the Landsat Global Archive Consolidation (LGAC) activity the PCD are incomplete or missing altogether. In order for scenes to be geometrically corrected to enable "stacking" for time series analyses, a geodetic accuracy threshold of less than 11.9m root mean square error (RMSE) based on a post-fit analysis of the data relative to the Global Land Survey 2000 (GLS2000) ground control. When using this metric in the context of post model fit radial RMSE relative to the GLS ground control, more than 57% of all OLI, 73% of all ETM+, and more than 60% of all TM scenes can be processed to L1T and satisfy this criteria. OLI data can be processed L1GT and meet these specifications because global positioning system (GPS) data are included in the ephemeris. However, for many scenes the reference grid is still off by up to 50 meters, so although the L1GTs may have better absolute accuracy, they will not stack with L1T data that are registered to the reference grid. Landsat ETM+ data with 70% or less cloud cover have a 82% success rate or better of being processed to L1T; Landsat TM data having 40% or less cloud cover have an 81% success rate or better of being processed to L1T.

There are three modes of near-real time (NRT) data collection and processing that may not immediately satisfy the collection definition criteria, two of which apply to Landsat 7 ETM+ data and the other applies to Landsat 8 OLI/TIRS: there may be a delay in generating ETM+ Level-1 products of sufficient geometric accuracy due to the need to update the bumper mode calibration coefficients; ETM+ products generated using

1

USGS

Collection Definition Progress

- **The USGS defined three basic categories of products**
  - **NRT (Near-real time) – products that are processed using ancillary data such as predicted ephemeris or bumper mode parameters that may be improved by reprocessing**
  - **Tier 1 – products that meet the criteria for the collection definition (i.e. enable time-series stacking, <11.9m RMSEr)**
  - **Tier 2 – products that do not meet the criteria for the collection definition and have been processed using the best known ancillary data**

A single collection (i.e. "collection 1") for all sensors (excluding MSS) as opposed to a separate collection per sensor

USGS

Collection Definition Study Findings Summary

- **Radiometric variability is not a factor**
  - **Operational land imager (OLI) temporal uncertainty is better than 0.3% on average**
    - **Based on on-board calibrator**
  - **ETM+ and TM are better than 2%**
    - **Based on top of atmosphere reflectance measured over pseudo-invariant calibration sites**
- **Geodetic accuracy vary by sensor, source data type, the quality of PCD, and level to which the data have been processed (L1T, L1GT, L1G)**
  - **Some data (e.g. TMA and LGAC) are of lower quality due to poor quality or missing PCD**

MSS is highly variable and is a separate discussion at this meeting

USGS

Near-real Time (NRT) Products

- **Three Modes of NRT data collection that do not satisfy geometric accuracy**
  1. **ETM+ products generated using predicted ephemeris data**
  2. **ETM+ products generated that utilize predicted bumper mode calibration coefficients**
  3. **OLI_TIRS products that utilize a preliminary line of sight (LOS) model based on "estimated" position of the scene select mirror.**
     - **TIRS coefficients will initially be updated periodically, initially quarterly, increasing in frequency to twice per month after sufficient telemetry has been collected**

- **These products will be made available for download as soon as they are processed and will roll off the download cache once they become Tier-1 or Tier-2 collection products**

USGS

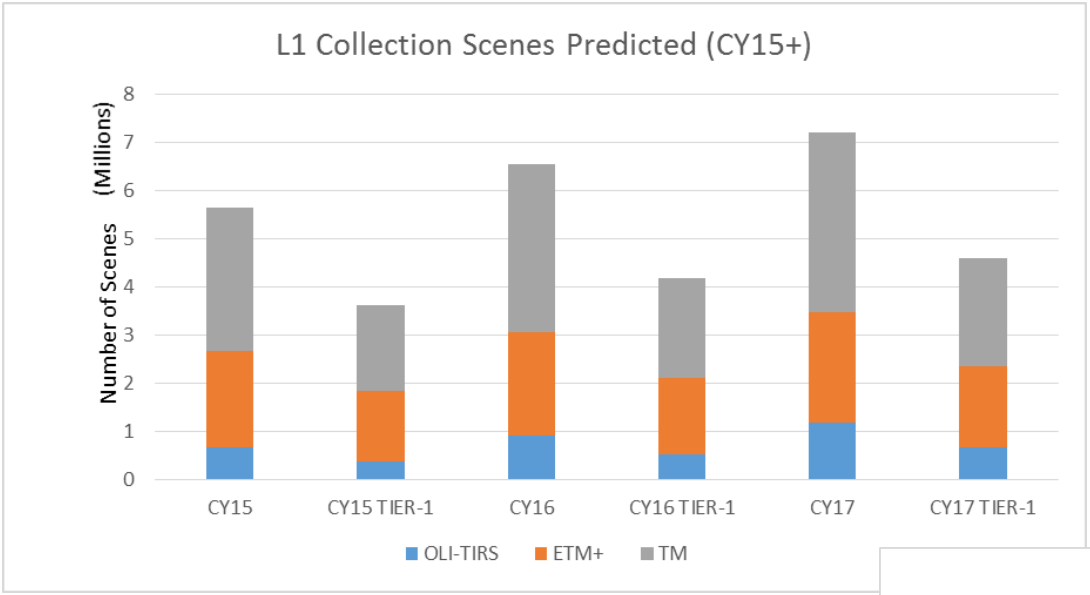
Tier-1 Collection Definition Summary

- **Need to be geometrically corrected to enable multi-spectral time-series stacking**
  - **Geodetic accuracy threshold of less than 11.9m radial root mean square error (RMSEr) based on a post fit analysis of data relative to the Global Land Survey (GLS) 2000 ground control**
  - **Results in about**
    - **57% of OLI_TIRS**
      - **A higher percentage of OLI science are collected over areas without ground control (Antarctic, Coastal Areas, Islands, higher cloud cover)**
      - **L1GTs can't perform post-fit verification to GLS so they are part of the TIER-2 category**
    - **73% of all ETM+**
    - **60% of all TM**

The intent is to make the full Tier-1 collection available for immediate download

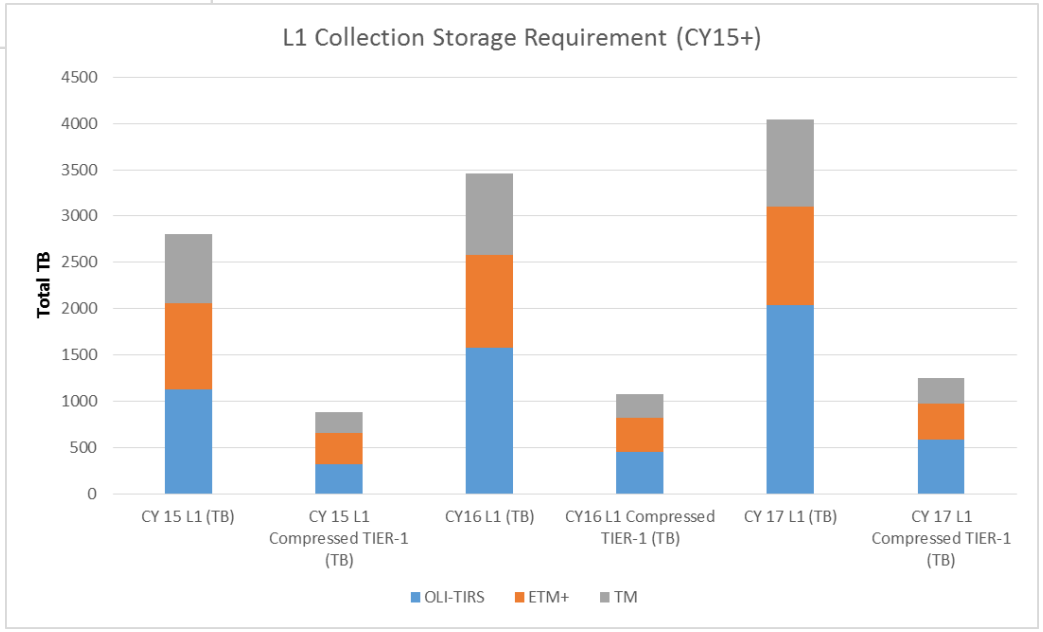
USGS

Collection Definition Product Estimates



TIER-1 Collection

Sensor	% Stackable
OLI-TIRS	57.00%
ETM+	73.00%
TM	60.00%



Tier-2 Product

- **Products that do not meet the definition of the collection and have been processed using the best known ancillary data**
  - **Definitive ephemeris**
  - **Latest bumper mode parameters**
  - **Latest TIRS LOS coefficients**

- **Data will be made available at the highest processing level through the Tier-2 collection or on-demand processing request**

USGS

Collection Identification

Product ID / File Name Convention

- **Current Proposed Product ID:**
  - **LXSS_LLL_PPPRRR_YYYYDDMM_yyyymmdd_CC_O**
    - **L = Landsat (constant)**
    - **X = Sensor (C = OLI/TIRS, E = ETM, T = TM, etc.)**
    - **SS = Satellite (e.g., 09 for Landsat 9, 10 for Landsat 10)**
    - **LLL = Processing level (L1T, L1G, L1S)**
    - **PPP = WRS path**
    - **RRR = WRS row**
    - **YYYYMMDD = Acquisition Year (YYYY) / Month (MM) / Day (DD)**
    - **yyyymmdd = Processing Year (yyyy) / Month (mm) / Day (dd)**
    - **CC = Collection number (e.g., 02)**
    - **O = "Optional" _R for Real-time or _N for Non-Collection**

    Example:

    LE07_L1T_029030_20140715_20140805_02

USGS

Initial International Cooperator Feedback Received

1. Human readable file naming becomes less important as systems handle the data in more automated ways, allowing for shorter file names
2. Removing the ground station ID was a point of concern for ICs
   - Only being proposed to be done to the Level 1 products and the raw/L0 data is unaffected and will still include the ground station ID
3. Minor collection # (or version) to the file name was discussed
   - Plan to use the processing date to provide that differentiator
   - Follows experiences from MODIS
4. The concept of adding time to the file name
   - It is not in the USGS concept to reprocess multiple times in a single day. If USGS does, it would likely be because of an error and the data would be replaced
5. Converging with a more industry-adopted processing level naming convention
   - There would be significant impacts if the USGS would implement this at this time
     - Naming conventions are entrenched in many user's systems and culture
   - A crosswalk document could be provided to show the USGS Landsat processing level naming conventions to those adopted by NASA and CEOS, for example

USGS

Considered the Sentinel 2 File Naming Convention

- **See the url:**
  https://earthobservation.wordpress.com/2015/12/18/my-first-experiences-with-sentinel-2a-data/

  - In summary users stated they prefer concise file names with information on the sensor, where it's from, and the day that the data was collected

More discussion during the Sentinel 2 discussion topic that John will be leading

USGS

Collection Data Management - Processing

- **Collection production for TM & ETM+ will be ready in May 2016**
- **L8 OLI collection processing will be ready in October 2016 (delayed due to TIRS)**
  - **Delayed due to TIRS alternative operations concept implementation**

Product Release / Activity	CY 2016				CY 2017			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
L4-7 Collection 1 Production		~May (1 yr)						
L8 TIRS new OPS Con Development		~Apr						
L8 Collection Development								
L8 Collection 1 Production				~Oct - Dec				

- **Original collection processing order**
  1. **Process OLI_TIRS and ETM+ data back to L8 OPS (May 2013)**
  2. **All ETM+, TM US data**
  3. **Rest of ETM+ / TM**

- **Planned collection processing order**
  1. **ETM+/TM for US**
  2. **Rest of ETM+/TM**
  3. **All OLI_TIRS when ready**

Collection Data Management - Distribution

- **While populating the collection, continue to offer the existing data through downloads and on-demand as it exists today**
- **Open the collection for distribution shortly after production commences?**
  - **Ensure new Tier-1 and select Tier-2 collection data is available for immediate download (cached on-disk)**
    - **Less demanded Tier-2 data will be available from tape (not immediate download)**
    - **Plan to persist the OLI_TIRS Tier-2 data on disk cache**
  - **Since the collection is being produced in the order published (previous slide) we don't recommend allowing on-demand for collection data**
    - **On-demand can greatly hinder production performance and reduce production efficiency**
- **Advantages**
  - **Allows users to continue to get the existing data while collection is being produced**
  - **Allows users to begin to download the new collection as it is being produced**
    - **Favorable for big-data users**
    - **Lightens burst load on the wide area network (WAN)**
- **Disadvantage**
  - **Increased storage requirements**
    - **Mitigated – Manageable through existing 1.6PB storage configuration of distribution cache and use of tape for low-demand data**

USGS

Collection Management Control

- **Traditionally we managed product updates through system releases**
- **Through feedback from the product roadmap implementation plans and the need for a more product focus especially given collection management, we've modified processes to focus on "Product Control"**
  - **One of the key goals is to ensure that the USGS is open with our plans / planning process and to gather feedback from multiple parties**
- **The Landsat Scientist will lead the product control to ensure**
  - **The product roadmaps for USGS are defined and seek science feedback**
  - **Feedback is solicited from the science team for product improvements and future product roadmaps**
  - **Product improvement priorities are solicited and defined**
  - **Communication among teams**
  - **Proper product documentation is published / available**

USGS

Communications Approach

- **Collection Definition Document and Collection Management Plan to be posted on Landsat Mission Web Site**
- **White papers to be synthesized for journal publication**
- **Announcements to be published referencing online documents**
  - **Landsat Update**
  - **Landsat Mission Headlines**
  - **Social Media**
  - **USGS and NASA Communications & Outreach**
  - **Notices posted on EarthExplorer & GloVis**
- **Targeted Emails**
  - **Bulk users**
  - **LST, LP DAAC UWG, DOIRSWG, AmericaView, etc.**
- **Other recommendations?**

USGS

Questions?